单样本率比较（单组目标值法）的样本量计算确切概率法及 R 语言实现

曾治宇
波科国际医疗贸易（上海）有限公司医学部 100020

[摘要] 单样本率比较（单组目标值法）的样本量计算常见的方法为正态近似法，有时伴相应的数据转换如平方根反正弦转换，而确切概率法通常需要商业统计软件寻值或编程实现。本文利用免费软件 R 语言编程实现单样本率比较确切概率法计算样本量，并且考虑到了确切概率法计算时检验效能与样本量非单调递增的关系，直接给出计算结果，也可作图直观显示检验效能与样本量的关系，希望能有助于这类研究的有效开展。
[关键词] 单组目标值；单样本率比较；样本量计算；确切概率法；R 语言

Exact probability sample size calculation for single proportion comparison (performance goal) with R language

Zeng, Zhiyu
Boston Scientific Corporation BSC Int'l Medical Trading (Shanghai) Co., Ltd. Beijing, 100020, China

[Abstract] The common method of sample size calculation for single proportion comparison (performance goal) is normal asymptotic approach, sometimes with corresponding data transformation such as squared arcsine, while exact probability usually needs commercial statistics software or programming. We use the free software R to calculate the sample size for single proportion via exact probability, and considering of the non-monotone increasing relationship between power and sample size with exact probability, we provide intuitive figure demonstration besides giving direct calculation results. We hope this will facilitate study design with performance goal.
[Key words] performance goal; single proportion comparison; sample size calculation; exact probability; R language

当前的临床研究设计中单组目标值法（performance goal, PG）的应用越来越广泛。对于计数资料的单组目标值法，从统计学角度来看其实就是单样本率的比较，其样本量计算常见的为正态近似法及平方根反正弦转换法。如果要进行确切概率计算，实现起来仍有些麻烦。国内已有作者利用 SAS[1, 2]或 Stata[3]软件编程计算，有些也考虑到了确切概率法计算时检验效能与样本量的非单调递增关系[2, 4]。本文介绍单样本率比较的样本量计算确切概率法的 R 语言（3.4.0 版本）实现。

原理与 R 语言编程

单样本率比较确切概率法基本思路为基于累计二项分布计算检验效能，参考文献给出的计算公式[1]，我们编写 R 语言程序如下：

```
#首先建立函数 exactpow 计算检验效能
exactpow<-function(side,alpha,p0,p1,n){
  if(side==1&p1>p0){
    k<-n
    while (1-pbinom(k,n,p0)<=alpha){  #pbinom 为 R 语言函数，计算二项分布的概率
      k=k-1
    }
    power=1-pbinom(k+1,n,p1)
  }
  if(side==1&p1<p0){
    k<-0
    while (pbinom(k,n,p0)<=alpha){
```

```
            k=k+1
        }
        power=pbinom(k-1,n,p1)
    }
    if(side==2){
        k1<-0
        while (pbinom(k1,n,p0)<=alpha/2){
            k1=k1+1
        }
        k2<-n
        while (1-(pbinom(k2,n,p0))<=alpha/2){
            k2=k2-1
        }
        power=pbinom(k1-1,n,p1)+1-pbinom(k2+1,n,p1)
    }
    return(power)
}
#然后根据检验效能，建立函数 exactn 求样本量
exactn<-function(side,alpha,beta,p0,p1,i){
    power<-1
    while (power>1-beta){
        power<-exactpow(side,alpha,p0,p1,i)
        i=i-1
    }
    return(i+2)
}
```

注意，单样本率比较的样本量计算，检验效能随样本量的增加呈现锯齿状增加[2, 4]，这里 R 语言编程的思路是从一个较大的样本量开始（比如 i 取值 2000），向下寻值直到检验效能低于设定值（比如 0.8），然后再返回上一个值，即为所求样本量。

实例 1

该例来自文献[1]，为验证某一治疗肝癌的组合治疗方案是否有效，拟进行临床试验设计。根据以往研究数据获知，肝癌的 5 年生存率为 50%，研究者预期新的组合治疗方案能使肝癌的 5 年生存率提高至 60%，试按照检验效能为 80%、检验水准为 0.05 的双侧检验估计本试验所需样本量。原文利用 nQuery 软件，在相应的样本量 n 一行反复尝试填入不同数据，直至获得检验效能达到或超过 80%即为所求样本量最终结果是 208 例。原文又用 SAS 编程计算，将 n=207 和 n=208 分别带入程序，得到的检验效能分别为 0.79110 和 0.81396。

利用 R 语言计算如下：

```
> exactn(side=2,alpha=0.05,beta=0.2,p0=0.5,p1=0.6,i=2000)
[1] 210
```

结果为 210，与原文略有差异。我们利用 R 语言计算样本量为 207，208，209 和 210 时的检验效能：

```
> exactpow(side=2,alpha=0.05,p0=0.5,p1=0.6,n=207)
[1] 0.791098
> exactpow(side=2,alpha=0.05,p0=0.5,p1=0.6,n=208)
[1] 0.8139611
> exactpow(side=2,alpha=0.05,p0=0.5,p1=0.6,n=209)
[1] 0.7979761
> exactpow(side=2,alpha=0.05,p0=0.5,p1=0.6,n=210)
[1] 0.8202484
```

可见样本量 207 和 208 时的检验效能与原文 SAS 寻值的结果完全一样，而样本量增加至 209 时检验效能又跌至 0.8 以下，待样本量增加至 210 时，检验效能升至 0.8 以上。利用 PASS （版本 11）及 G*Power（版本 3.1）软件进行验算，与我们的结果一致，看来手动寻值法还是容易出错。我们继续利用 R 语言作图直观地看一下，样本量 199 时检验效能首次达到 0.8，而样本量 210 及以上时，检验效能稳定在 0.8 以上（图 1）。

```
j<-1;x<-c(0);y<-c(0)
for (i in c(188:222)){
   x[j]<-i
   y[j]<-exactpow(2,0.05,0.5,0.6,i)
   j<-j+1
}
plot(x,y,type="l",xlab="样本量",ylab="检验效能")
abline(h=0.8,lty=3)
abline(v=210,lty=3)
```

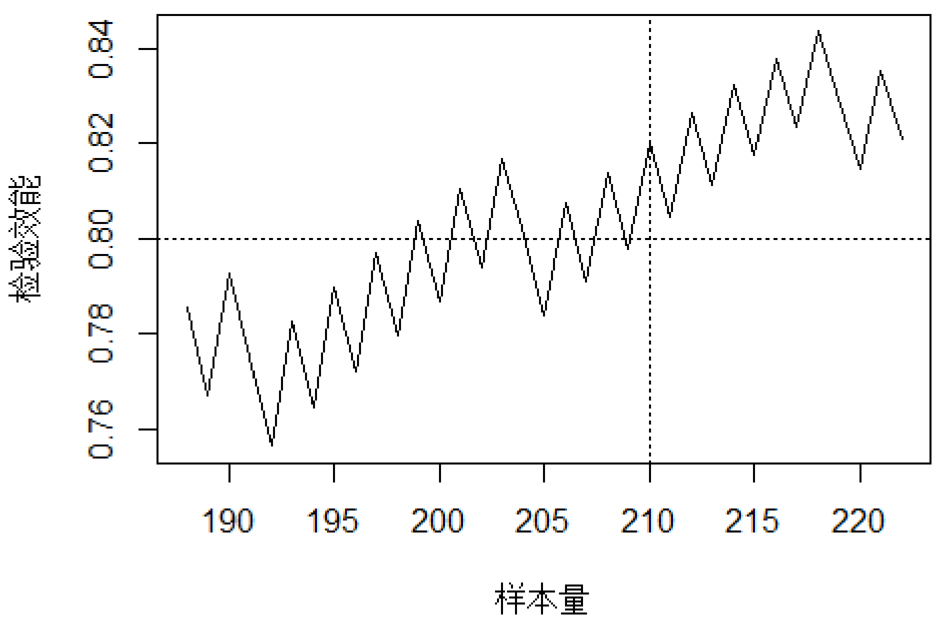


图 1 样本量与检验效能

实例 2

再看一个预期率小于目标值的例子，如设定 $\alpha=0.05$，已知总体率 $\pi_0=0.07$，预期的总体率 $\pi_1=0.03$，双侧检验，80%的检验效能。原文进行 SAS 编程计算，考虑到检验效能的非单调递增，最终样本量计算结果为 277[2]。R 语言计算如下：

```
> exactn(2,0.05,0.2,0.07,0.03,2000)
[1] 277
```

讨论

本文介绍了单样本率比较的样本量计算确切概率法的 R 语言实现，该方法考虑到了确切概率法计算时检验效能与样本量非单调递增的关系，直接给出计算结果，也可作图直观显示检验效能与样本量的关系，所有计算均通过 PASS 及 G*Power 软件验证。

国内已有作者利用 SAS[1, 2]或 Stata[3]软件编程计算，有些也考虑到了确切概率法计算时检验效能与样本量的非单调递增关系[2, 4]，我们的方法在计算确切概率时与前者是一致的，但寻值的思路是从一个较大的样本量开始，向下寻值直到检验效能低于设定值（比如 0.8），然后再返回上一个值，即为所求样本量，这样在编程上显得更为简洁，但缺点是计算量可能会较大，好在对于现代计算机来说，这并不是个问题。

文献介绍的利用样本量计算软件 nQuery 需要手动寻值查找样本量，容易出错[1]。而

PASS 和 G*Power 软件可以较为方便地通过作图找到合适的样本量。我们利用 R 语言编程，直接给出所需的样本量，并且 R 语言是免费软件，适用范围会更广泛。

希望本文能有助于这类研究在国内的有效开展。

参考文献

[1] 唐欣然，张惠风，揭著业，等. 样本量估计及其在 nQuery 和 SAS 软件上的实现——率的比较(一)[J]. 中国卫生统计,2012,29(5).

[2] 刘江美，陈平雁. 单样本率确切概率检验的样本量与检验效能非单调变化关系的研究[J]. 中国卫生统计,2012,29(2).

[3] 肖林海，赵耐青. 单样本率精确概率检验的样本量估计方法及在 Stata 中的实现[J]. 中国卫生统计,2014,31(4).

[4] 曾治宇，林娜，张明东，等. 单样本率比较（单组目标值法）的样本量计算及其简便实现 in press[J]. 中国卫生统计,2017